Profiling MOOC course returners

Citation for published version:
Kovanovic, V, Joksimovic, S, Gasevic, D, Owers, J, Scott, A-M & Woodgate, A 2016, Profiling MOOC course returners: How does student behavior change between two course enrollments? in *Proceedings of the 3rd Annual ACM Conference on Learning at Scale (L@S 2016).* ACM Press, pp. 269-272. https://doi.org/10.1145/2876034.2893431

Digital Object Identifier (DOI):
10.1145/2876034.2893431

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Proceedings of the 3rd Annual ACM Conference on Learning at Scale (L@S 2016)

Profiling MOOC Course Returners: How Does Student Behavior Change Between Two Course Enrollments?

Vitomir Kovanović
School of Informatics,
University of Edinburgh
Edinburgh, United Kingdom
v.kovanovic@ed.ac.uk

Srećko Joksimović
Moray House School of Education,
University of Edinburgh
Edinburgh, United Kingdom
s.joksimovic@ed.ac.uk

Dragan Gašević
Moray House School of Education
and School of Informatics,
University of Edinburgh
Edinburgh, United Kingdom
dgasevic@acm.org

James Owers
School of Informatics,
University of Edinburgh
Edinburgh, United Kingdom
james.owers@ed.ac.uk

Anne-Marie Scott
Information Services,
University of Edinburgh
Edinburgh, United Kingdom
anne-marie.scott@ed.ac.uk

Amy Woodgate
Information Services,
University of Edinburgh
Edinburgh, United Kingdom
amy.woodgate@ed.ac.uk

L@S 2016, April 25-26, 2016, Edinburgh, Scotland UK.
ACM 978-1-4503-3726-7/16/04.
http://dx.doi.org/10.1145/2876034.2893431

Abstract

Massive Open Online Courses represent a fertile ground for examining student behavior. However, due to their openness MOOC attract a diverse body of students, for the most part, unknown to the course instructors. However, a certain number of students enroll in the same course multiple times, and there are records of their previous learning activities which might provide some useful information to course organizers before the start of the course. In this study, we examined how student behavior changes between subsequent course offerings. We identified profiles of returning students and also interesting changes in their behavior between two enrollments to the same course. Results and their implications are further discussed.

Author Keywords

MOOCs; Clustering; Learning Analytics; Student Behavior; Self-regulated learning; Educational technology use;

ACM Classification Keywords

H.3.3 Information Search and Retrieval: Clustering; K.3.1 Computer Uses in Education: Distance learning; I.6.4 Model Validation and Analysis; I.6.5 Model Development; I.5.3 Clustering;

#	Course	Offers
1	Artificial Intelligence Planning	1, 2, 3
2	Animal Behavior and Welfare	1, 2
3	AstroTech: The Science and Technology behind Astronomical Discovery	1, 2
4	Astrobiology and the Search for Extraterrestrial Life	1, 2
5	The Clinical Psychology of Children and Young People	1, 2
6	Critical Thinking in Global Challenges	1, 2, 3
7	E-learning and Digital Cultures	1, 2, 3
8	EDIVET: Do you have what it takes to be a veterinarian?	1, 2
9	Equine Nutrition	1, 2, 3
10	Introduction to Philosophy	1, 2, 3, 4
11	Warhol	1, 2

Table 1: Courses used in this study.

Introduction

One of the important characteristics of Massive Open Online Courses (MOOCs) is their complete openness to students with different learning goals, motivations, and backgrounds [2]. As such, instructors have typically very little or no information who students who enrolled in their courses are. Typically, the primary sources of information about registered students are MOOC platform demographics and pre-course surveys which are often filled out by only a small subset of all course registrants.

Besides demographic and survey data, another valuable source of information about students enrolling a particular MOOC could be obtained from their previous enrollments. With the growing number of MOOCs offered and with their multiple instances, it became common to have the same students enrolled in different MOOCs from the same institution and even enrolled to different offerings of the same course. This data could provide valuable information about students before the start of the course and can help course designers and instructors to better cater their online courses to the target population.

Not only can data from previous course offerings be used to improve course offerings, but it can also be used to study learning similarly to the repeated measurement experiments. As the selection of MOOC participants is out of the instructor's control, analysis of the data from several offerings of the same course can provide some insight into the choices student make regarding the use of the available resources, tools, and affordances.

Research Questions

Given the potentials of the MOOC data to understand learning behavior and choices about the use of available resources, tools, and technologies, we examined MOOC trace data from the students who enrolled in the same course at least twice. The primary research questions addressed in this paper are:

RQ1: *What are common behavioral profiles of students who enroll MOOCs multiple times?* This question is the first step in our analysis and it allows for examining whether there are any particular forms of MOOC engagement by the students who enroll in the same courses multiple times.

RQ2: *How do students change their behavior between subsequent offerings of the same course?* This follow-up question is a natural extension to our first question and focuses on student self-regulation of learning. Do students change something in their behavior between two offerings or they simply continue with the same form of participation as they did the first time?

Method

Dataset

The data for this study comes from the 28 offerings of the 11 different MOOCs offered by the University of Edinburgh on the Coursera platform (Table 1). In our analysis, we examined only data about students' first and second enrollment. That is, we did not analyze students who enrolled only once, and we also excluded any subsequent (i.e., third or fourth) enrollments. In total, we had 26,025 double course enrollment records (52,050 course enrollment records).

Variable	Description
Days	No. of days active
Sub.	No. of submitted assignments
Wiki	No. of wiki page views
Disc.	No. of discussion views
Posts	No. of discussion messages written
Quiz.	No. of quizzes attempted
Quiz. Uni.	No. of different quizzes attempted
Vid. Uni.	No. of different videos watched
Vid.	No. of videos watched

Table 2: Clustering variables.

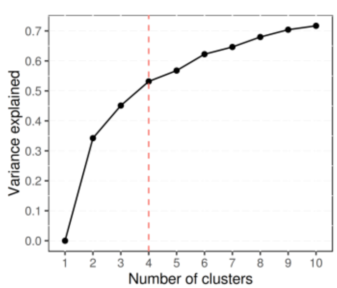

Figure 1: Variance explained by a given number of clusters

Analysis procedure

To answer our research questions, we conducted a cluster analysis of the 52,050 enrollment records using the variables listed in Table 2. As overall student activity in each course was slightly different, we first performed unitization with zero minimum (i.e., x-min/range) per each course offering, along with a z-score normalization on the whole corpus. That way we ensured that 1) specifics of each course (and each course offering) were taken into account for scaling each classification variable, and 2) all variables were on the same scale to ensure the equal importance of variables. Given that there is also a large number of students who only enrolled in courses (and have not accessed them afterward), we removed those records from our subsequent cluster analysis and assigned them a predefined "Enroll Only" cluster.

We performed K-means clustering using Lloyd's algorithm (with ten restarts and a maximum of 300 iterations) for values of K between 2 and 10 and the evaluated the percentage of variance explained by the different clustering solutions (Figure 1). We selected the K-means algorithm as the size of our dataset (52,050) was too large for analysis using some of the more sophisticated classification techniques that involve pairwise distance matrix. Finally, after identifying student clusters, we examined a transition graph between students' first and second enrollment to see what the most common cluster transitions are.

Results and Discussion

Clustering

Our results and scree plot analysis (Figure 1) revealed four clusters of student behavior in our dataset, in addition to our "Enroll Only" cluster of students who did not exhibit any course activity. In total, we identified five clusters of the behavior of returning students. The cluster centers are shown in Figure 2 while their relative sizes are shown in Table 3. These results are aligned with the previous work on online courses [3] and MOOCs [1] that showed similar disproportions between highly active and inactive students. The identified clusters (Figure 2) reveal that the largest part (85% of all enrolled students) have no or have very little course activities. Around 10% of the students focused primarily on viewing video lectures, while 4.1% of students were highly engaged and, besides watching videos, also utilized quizzes and engaged in homework assignments. Finally, less than 1% of student put an emphasis on online discussions, while being less engaged with video lectures. This cluster of students also stayed longest active in courses.

Cluster transitions

To investigate how student behavior changes between subsequent course enrollments, we constructed a directed state transition graph (Figure 3) which shows what percentage of first enrollment cluster members transferred to other clusters (or remained within the same cluster). The majority of students from all the clusters except the "Social" cluster either just enrolled in a course or had very low level of engagement. A certain number of students who utilized both video lectures and quizzes during their first enrollment either retained the same level of engagement or focused primarily on video lectures in the second course enrollment. These two patterns are likely driven by the goal of obtaining course certificate or brushing up on a particular course topic. Finally, the most interesting finding is related to the students from the "Social" cluster who had the highest level of participation in

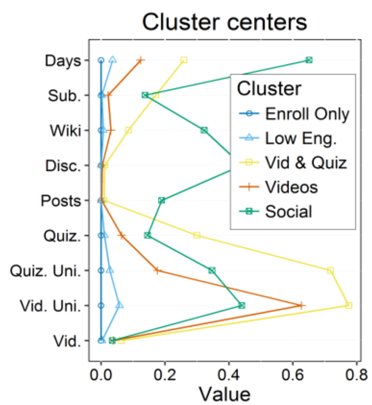

Figure 2: Cluster centers

Cluster	Students	%
Enroll Only (E)	22,932	44.1
Low Engagement (LE)	21,776	41.8
Videos & Quizzes (VQ)	2,120	4.1
Videos (V)	5,128	9.9
Social (S)	94	0.2

Table 3: Cluster sizes.

	Second enrollment				
First enrollment	E	LE	VQ	V	S
E	.65	.28	.03	.04	.00
LE	.43	.45	.04	.08	.00
VQ	.35	.37	.16	.11	.00
V	.35	.42	.08	.14	.00
S	.21	.26	.11	.14	.28

Table 4: Cluster transitions as percentage of first enrollments.

online discussions and also most days spent in the course. While a certain number of students became disengaged in the next offer of the course, a large chunk of them (28%) kept their level of participation, signaling the goal of engaging with other learners rather than the prescribed course content.

Implications and Future Work

There are several practical implications of our findings. First, as a majority of students who were not active (or had low levels of activity) in their first enrollment were likely to stay inactive, course instructors might consider targeting those particular students with a certain set of instructional interventions which would increase their levels of participation. Similarly, students who exhibited high levels of activity in the first offer might be targeted with interventions that would encourage them to participate more in the discussions, or with interventions related to particularly challenging course content (as indicated by their quiz and assignment scores in the previous enrollment). Finally, through identification of socially engaged students, instructors might identify suitable community teaching assistants which could be then better supported by the instructional team.

Although our analysis provides interesting insights, there are still many potential areas for the future research. In particular, understanding the relationship between course participation and answers to pre-course surveys and certificate earning in the first enrollment. By answering of these important questions, we aim to enable instructors to better cater their online courses to the prospective students, and also to better understand MOOC learning in general.

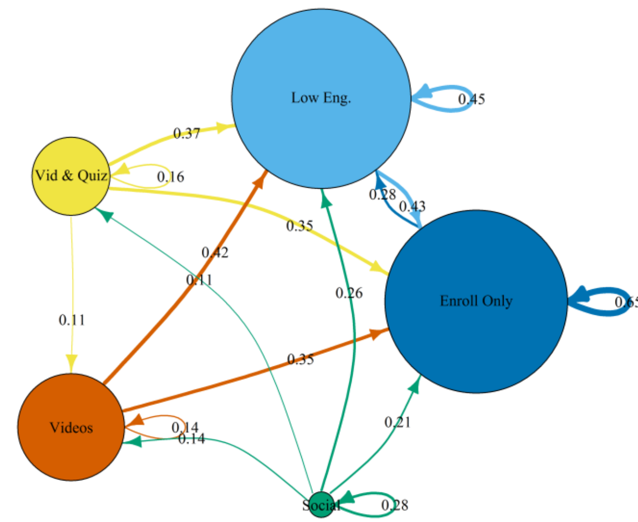

Figure 3: Cluster transition graph. Sizes of nodes represent the number of students while edge labels represent percentage of source cluster transitioning to the destination cluster

References

1. Doug Clow. 2013. MOOCs and the Funnel of Participation. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, ACM, 185–189.
2. R.F.a Kizilcec and E.b Schneider. 2015. Motivation as a lens to understand online learners: Toward data-driven design with the OLEI scale. *ACM Transactions on Computer-Human Interaction* 22, 2.
3. Vitomir Kovanović, Dragan Gašević, Srećko Joksimović, Marek Hatala, and Olusola Adesope. 2015. Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions. *The Internet and Higher Education* 27: 74–89.